Machien Learning and Inductive Inference [H02C1a]

Xinhai Zou (r0727971)

Contents

1 Lecture 1: Introduction, Version spaces

1.1 Some ML examples in practice

1. Autonomous cars
2. The Robosail project
3. The Robot Scientist
4. Infra Watch, ”Hoolandse brug” - the bridge
5. Language learning
6. Automating manual tasks

1.2 Machine Learning

Definition of machine learning: it is the study of how to make programs improve their performance on certain tasks from own (experience).
In this case:

- ”performance” = speed, accuracy
- ”experience” = earlier observations

Machine Learning vs. other AI
In **machine learning**, the key is **data**, examples of questions and their answer; observations of earlier attempts to solve the problem
In **inductive inference**, it is reasoning from **specific** to **general**, statistics: sample -> population; from **concrete observations** -> **general theory**

1.3 Machine Learning learning landscape

- tasks
  - clustering
  - classification
  - regression
  - reinforcement learning
- techniques
  - Convex optimization
  - Matrix factorization
  - Transfer learning
  - Learning theory
  - Greedy search

- models
  - automata
  - neural network
  - deep learning
  - statistical relational learning
  - decision trees
  - support vector machines
  - nearest neighbors
  - rule learners
  - bayesian learning
  - probabilisitc graphical models
- applications
  - natural language processing
  - vision
  - speech
- related courses
  - neural computing
  - support vector machine
  - uncertainty in AI
  - data mining
  - genetic algorithms and evolutionary computing

1.4 Some basic concepts and terminology

- Predictive learning
  - Definition: learn a model that can predict a particular property/ attribute/ variable from inputs
  - Binary classification: distinguish instances of class C from other instances
  - Classification: assign a class C (from a given set of classes) to an instances
  - Regression: assign a numerical value to an instance
  - multi-label classification: assign a set of labels (from a given set) to an instance
  - multivariate regression: assign a vector of numbers to an instances

- multi-target prediction: assign a vector of values (numerical, categorical) to an instances
- Descriptive learning
  - Definition: given a dataset, describe certain patterns in the dataset, or in the population it is drawn from
- Typical tasks in ML
  - function learning: learn a function X->Y taht fits the given data
  - distribution learning: distribution learning
    - parametric: the function family of the distribution is known, we only need to estimate its parameters
    - non-parametric: no specific function family assumed
    - generative: generate new instances by random sampleing from it
    - discriminative: conditional probability distribution
- Explainable AI (XAI)
  - Definition: means that the decisions of an AI system can be explained
  - Two different levels:
    - We understand the (learned) model
    - We understand the individual decision

1.5 Input formats (predictive learning)

- Set
  - training set: a set of examples, instance descriptions that include the target property (a.k.a. labeled instances)
  - prediction set: a set of instance descriptions that do not include the target property ('unlabeled' instances)
  - prediction task: predict the label of the unlabeled instances
- Outcome of learning process
  - transductive learning: the predictions themselves
  - inductive learning: a function that can predict the label of any unlabeled instance
- Explainable AI
  - interpretable: can be intepred
  - black-box: non-interpretable
- Learning

- Supervised learning: from labeled
    - Unsupervised learning: from unlabeled
    - Semi-supervised learning: from a few labeled and many unlabeled

- Format of input data
    - input is often assumed to be a set of instances that are all described using the same variables (features, attributes)
    - i.i.d.: independent and identically distributed
        * tabular data (NN)
        * sequences
        * trees
        * graph
        * raw data: learning meaningful feaures from raw data
        * knowledge: inductive logic programming

1.6 Output formats, methods (predictive learning)

The **output** of a learning system is a model.

- output
    - parametrized functions
    - ocnjunctive concepts: a conjuntive concept is expressed as a set of conditions, all of which must be true
    - rule sets (if...then...else...)
    - decision trees
    - neural networks
    - probabilisitc graphical models
- search methods
    - discrete spaces - methods: hill-climbing, best-first
    - continuous spaces - methods: gradient descent
- typically
    - model structure not fixed in advanced - discrete
    - fixed model structure, tune numerical parameters - continuous
- hypothesis space
    - definition: all possible instances
    - for robot example: {B,R,M,?} x {S,T,?} x {L,W,?} x {1,2,?}

- Verson space
  - using candidate elimination
  - pros
    - can be used for discrete hypothesis spaces
    - search for all solutions, rather than just one, in an efficient manner
    - importance of generality ordering
  - cons
    - not robust to noise
    - only conjunctive concepts

2 Lecture 2: Induction of decision tree

2.1 Overview of DT

- A decision tree represents a decision procedure where
  - you start with one question
  - the answer will determine the next question
  - and repeat, untill you reach a decision
- We will usually call the questions "tests" and the decision a "prediction"
- attribute
  - input attribute X = {X_1, X_2 ..., X_n}
  - target attribute Y
  - the tree represents a function f: X -> Y
- Example: Playing Tennis Tree
  - Outlook: X_1 = {Sunny, Overcast, Rainy}
  - Humidity: X_2 = {High, Normal}
  - Wind: X_3 = {Strong, Weak}
  - Tennis: Y = {Yes, No}
  - The tree represents a function Outlook x Humidity x Wind -> Tennis
- Boolean tree
- Continuous input attributes
  - We cannot make a different child node for each possible value!
  - Solution: use comparative test -> a finite number of possible outcomes

- Type of trees
  - target attribute Y is nomial -> classification tree
  - target attribute Y is numerical -> regression tree
- **Advantages of Tree (Why tree?)**
  - Learning and using tree is **efficient**
  - Tend to have **good predictive accuracy**
  - Tree is **interpretable**

2.2 Learn trees from data

- Two tasks for DT
  - Task 1: find the smallest tree T such that ∀(x,f(x))∈D: T(x)=f(x) (meaning that only fullfill current data set)
  - Task 2: find the tree T such that for x drawn from population D, T(x) is (on average) maximally similar to f(x) (T:model tree from data set D, f(x):true function in population D)
    - loss function: l: Y_1 x Y_2 -> R (where Y_1 is predicted value, Y_2 is actual value)
    - risk R of T, the expectation of loss function, is $E_{x \sim D}[l(T(x), f(x))]$, which is needed to be minimal.
- the basic principle
  - The approach is known as ”Top-down induction of decision trees (TDIDT)”, or ”recursive partitioning”
    - 1. start with the full data set D
    - 2. find a test such that examples in D with the same outcome for the test tend to have the same value of Y
    - 3. split D into subsets, one for each outcome of that test
    - 4. repeat this procedure on each subset that is not yet sufficiently ”pure” (meaning, not all elements have the same Y)
    - 5. keep repeating until no further splits possible

3 Lecture 3: Learning sets of rules

4 Lecture 4: Instance-based learning, Clustering

5 Lecture 5: Evaluating hypotheses

6 Lecture 6: Numerical approaches (ANN, SVM), Computational learning theory

7 Lecture 7: Probabilistic approaches, Ensembles

8 Lecture 8: Reinforcement learning

9 Lecture 9-10: Inductive logic programming